DOCUMENT RESUME

ED 032 224

SE 007 486

Annual Report to the National Science Foundation on Contract NSF-C414 Task I, 1 June 1967 through 31 May 1968.

American Chemical Society, Columbus, Ohio. Chemical Abstracts Service.

Spons Agency-National Science Foundation, Washington, D.C.
Pub Date May 68
Note-14p.
EDRS Price MF-$0.24 HC-$0.80
Descriptors-Annual Reports, *Chemistry, Computer Storage Devices, Information Dissemination, *Information Processing, *Information Science

The work reported here was performed between 1 June 1967 and 31 May 1968 under Task I of the contract between Chemical Abstracts Service (CAS) and the National Science Foundation (NSF). The objective of this contract was to achieve an operational, integrated, man-machine system for manipulating information about chemical substances. The contract was directed at the development of a computer-based Chemical Compound Registry System through the buildup of machine and manual files of chemical structural representations, chemical nomenclature, molecular formulas, and bibliographic citations associated with the registere d compounds. To meet this objective the contract called for the registration of compounds from several sources. Progress was made in the number of compounds registered from all sources. CAS worked to extend the range of compounds that could be machine-registered and to improve the overall efficiency and effectiveness of the system by shifting additional tasks to the computer to improve the technical and/or economic performance of the system. (RR)

ED032224

F-NSF

ANNUAL REPORT

to the

NATIONAL SCIENCE FOUNDATION

on

CONTRACT NSF-C414 TASK I

1 June 1967 through 31 May 1968

CHEMICAL ABSTRACTS SERVICE

AMERICAN CHEMICAL SOCIETY

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

Columbus, Ohio

31 May 1968

SE 007 486

ANNUAL REPORT

to the

NATIONAL SCIENCE FOUNDATION

on

CONTRACT NSF-C414 TASK I

1 June 1967 through 31 May 1968

CHEMICAL ABSTRACTS SERVICE

AMERICAN CHEMICAL SOCIETY

Columbus, Ohio

31 May 1968

CONTENTS

		Page
I.	SUMMARY.	1
II.	REGISTRATION OF CHEMICAL SUBSTANCES.	2
III.	REGISTRY IMPROVEMENTS.	4
IV.	TABULAR DATA	6

I. SUMMARY

The work reported here was performed between 1 June 1967 and 31 May 1968 under Task I of Contract NSF-C414 between CAS and the National Science Foundation (NSF). The overall objective of this contract is to achieve an operational, integrated, man-machine system for manipulating information about chemical substances. This system must be capable of high speed, flexibility, and depth in responding to the information needs of those who use chemical information.

Task I of Contract NSF-C414 is directed at the development of a computer-based Chemical Compound Registry System through the buildup of machine and manual files of chemical structural representations, chemical nomenclature (including the systematic nomenclature developed for and used in Chemical Abstracts (CA) indexes, as well as trivial and other types of systematic names), molecular formulas, and bibliographic citations associated with the registered compounds.

To meet this objective the contract calls for the registration of compounds from several sources, including the current literature as indexed in CA, and the extensive CAS internal reference files. Progress has been made in the number of compounds registered this year from all sources. In addition, CAS has worked to extend the range of compounds that can be machine-registered and to improve the overall efficiency and effectiveness of the system by shifting additional tasks to the computer to improve the technical and/or economic performance of the system.

II. REGISTRATION OF CHEMICAL SUBSTANCES

To date, 887,104 unique substances have been filed in the system. These were obtained from CA indexes and CA reference files, as well as from other sources not included in Task I. The total number of registrations carried out is 1,697,257, representing the 887,104 unique substances on file plus 810,153 registrations of substances that matched one already on file.

A. Registration of Compounds Indexed in *Chemical Abstracts*

During the period 29 May 1967 through 31 May 1968, CAS added 248,051 unique compounds to the Registry Files as a result of registrations from the current literature as indexed in CA Subject Indexes. This brings to 693,749 the total number of compounds registered from CA index processing since Contract C414 began. Registration for Volumes 62-66 of CA has been completed at this time. Volume 67 has so far contributed 142,640 registry transactions which is an estimated 70% of the total to be included. Some 75% of the structures have been prepared, and 70% of the names have been keyboarded. Table III summarizes the machine and manual registrations resulting from registration of compounds from CA indexes.

B. CAS Reference Files

The registration of compounds from fourteen CAS reference files contributed 8194 unique structures to Registry Files during the year 29 May 1967 through 31 May 1968. This brings to a total 113,430 the number of

compounds registered from CAS reference files under Contract C414. Table III identifies the files and the number of compounds registered from each. Generally the contribution of any one file is relatively small, since only the updates to the file must now be registered, since registration of the basic files was completed during earlier work under the contract.

C. Sources Not Included in Task I

In addition to the CA indexes and the CA reference files, CAS has registered 18,923 unique compounds from sources not specified as part of Task I. These include other government-funded work such as Task III of Contract C414.

III. REGISTRY IMPROVEMENTS

A. Registry Redesign and Reprogramming

During this year, the redesign and reprogramming of basic programs of the Structure Registry was completed. The 360 Structure Registry programs were installed on 8 April 1968. All programs are working satisfactorily and only minor program problems were encountered at installation time. These were corrected rapidly and caused no processing delays.

Alterations have been made to the bibliography system so that all nomenclature can be retained on the Bibliography File for more economical retrieval via the Registry Number. Modifications were also made which allow the system to handle proposed structures through the retention of several molecular formulas.

B. Improved Textual Descriptors

CAS has developed improved conventions for deriving textual descriptors for certain specific types of alkaloids. The basis for the procedures is the structure and/or systematic name. The textual descriptors consist of the name of the parent structure with alphanumeric prefixes for positions whose stereochemical detail is not implied by the name. For example, in the alkaloid whose CA index name is Crinan-3A-ol,1B,2B-epoxy-7-methoxy, Crinan is the parent compound and nodes 1,2,3, have stereochemistry that must be specified. The special descriptor for this alkaloid is: 1B,2B,3A-CRINAN.

Additional capabilities for metallic compounds, text descriptors, and abnormal masses have been added to the connection table edit program.

C. Registration of Additional Classes of Compounds

Additional classes of compounds, i.e., inorganic compounds, coordination compounds, elements and their isotopes and allotropes, and compounds with more than 150 but less than 253 non-hydrogen atoms are now being registered on a routine basis.

IV. TABULAR DATA

TABLE I

Summary of Registration, 1 June 1965 - 31 May 1968

TABLE II

Total Number of Registrations by Source

TABLE III

Additions to the File of Unique Substances by Source

TABLE IV

First Source of Names on Nomenclature File

FIGURE I

Percentage of Registry Transactions that Result in Substances New to the Registry File

TABLE I

SUMMARY OF REGISTRATION
1 June 1965 through 31 May 1968

Source	Unique Substances Registered	
A - Current Literature		
CA Volume 62	114,816	
CA Volume 63	131,197	
CA Volume 64	119,869	
CA Volume 65	106,265	
CA Volume 66	98,336	
CA Volume 67	92,650	
Subtotal		663,133
B - CAS Reference Files		
Alkaloid File	697	
Colour Index	3,270	
Drug File	336	
Fluorine	42,157	
CA Formula Index Cross-References	953	
Lange Handbook of Chemistry	4,183	
Merck Index	6,005	
Pesticide Index	156	
CAS Reference File	17	
Ring Index (plus supplements)	21,901	
SOCMA Handbook	6,475	
CA Specific Volume Cross-References	1,723	
†Steroid File	349	
CAS Subject Index Cross-References	15,809	
†Terpene File	518	
USAN (United States Adopted Names)	82	
Subtotal		104,631
C - Task III		2,146
D - NON-C414		68,157
E - Manual Registration (Includes 6,097 mixtures)		49,037
F - Total		887,104

† Not routinely updated

TABLE II

TOTAL NUMBER OF REGISTRATIONS BY SOURCE

Source	Registrations Through 31 May 1968	
A - Handled By Machine and Manually		
1 - Current Literature		
CA Volume 62	166,851	
CA Volume 63	189,845	
CA Volume 64	173,317	
CA Volume 65	176,900	
CA Volume 66	159,580	
CA Volume 67	142,640[a]	
2 - CAS Reference Files		
Alkaloid File	2,987	
Colour Index	9,059	
Drug File	855	
Fluorine	59,634	
CA Formula Index Cross-References	2,809	
Lange Handbook of Chemistry	8,939	
Merck Index	16,538	
Pesticide Index	482	
CAS Reference File	47	
Ring Index (plus supplements)	28,393	
SOCMA Handbook	6,753	
CA Specific Volume Cross-References	4,938	
[b]Steroid File	1,070	
CAS Subject Index Cross-References	22,524	
[b]Terpene File	1,829	
USAN (United States Adopted Names)	828	
3 - Task III	20,620	
4 - NON-C414	113,567[c]	
Total Structure Registrations		1,310,405[d]
B - Handled By Name Match		
CA Volumes	156,237	
CAS Reference File	25,892	
Task III	18,732	
NON-C414	185,991	
Total Registrations by Registry Number	386,852	
Total		1,697,257[d]

a Processing incomplete as of 31 May 1968
b Not routinely updated
c Not charged to NSF
d Includes all compounds processed before 29 May 1965

TABLE III

ADDITIONS TO THE FILE OF UNIQUE SUBSTANCES BY SOURCE

Source	Through 29 May 1967	29 May 1967 to 31 May 1968	Total As Of 31 May 1968
A - Machine Registration			
1 - Current Literature			
CA Volume 62	109,021	5,795	114,816
CA Volume 63	125,498	5,699	131,197
CA Volume 64	113,374	6,495	119,869
CA Volume 65	87,787	18,478	106,265
CA Volume 66	6,309	92,027	98,336
CA Volume 67	-	92,650	92,650
2 - CAS Reference Files			
Alkaloid File	693	4	697
Colour Index	3,149	121	3,270
Drug File	303	33	336
Fluorine	42,147	10	42,157
CA Formula Index Cross-References	882	71	953
Lange Handbook of Chemistry	3,355	828	4,183
Merck Index	5,779	226	6,005
Pesticide Index	149	7	156
CAS Reference File	-	17	17
Ring Index (plus supplements)	20,034	1,867	21,901
SOCMA Handbook	6,414	61	6,475
CA Specific Volume Cross-References	1,625	98	1,723
†Steroid File	349	-	349
CAS Subject Index Cross-References	14,700	1,109	15,809
†Terpene File	518	-	518
USAN (United States Adopted Names)	37	45	82
3 - Task III	1,672	474	2,146
4 - NON-C414	52,572	15,585	68,157
Subtotal	596,367	241,700	838,067
B - Manual Registration			
CA Volume 62	804	5,828	6,632
CA Volume 63	471	2,738	3,209
CA Volume 64	1,176	1,928	3,104
CA Volume 65	1,258	7,388	8,646
CA Volume 66	-	6,312	6,312
CA Volume 67	-	2,713	2,713
CAS Reference Files	5,102	3,697	8,799
Task III	5,735	2,951	8,686
NON-C414	623	313	936
Subtotal	15,169	33,868	49,037
C - Grand Total	611,536	275,568	887,104

† Not routinely updated

TABLE IV

FIRST SOURCE OF NAMES ON NOMENCLATURE FILE

Source	Names on File 31 May 1968
A - Current Literature	
CA Volume 62	169,859
CA Volume 63	178,307
CA Volume 64	218,198
CA Volume 65	211,075
CA Volume 66	182,308
CA Volume 67	34,093[a]
B - CAS Reference Files	
Alkaloid File	3,142
Colour Index	38,363
Drug Index	2,186
Fluorine	51,491
CA Formula Index Cross-References	3,401
Lange Handbook of Chemistry	11,791
Merck Index	25,961
Pesticide Index	3,373
CAS Reference File	2,896
Ring Index (plus supplements)	19,844
SOCMA Handbook	27,292
CA Specific Volume Cross-References	13,424
[b]Steroid File	2,400
CAS Subject Index Cross-References	60,183
[b]Terpene File	3,181
USAN (United States Adopted Names)	2,421
Task III (NLM-FDA)	56,989
NON-C414	74,319[c]

[a] Processing incomplete as of 31 May 1968
[b] Not routinely updated
[c] Not charged to NSF

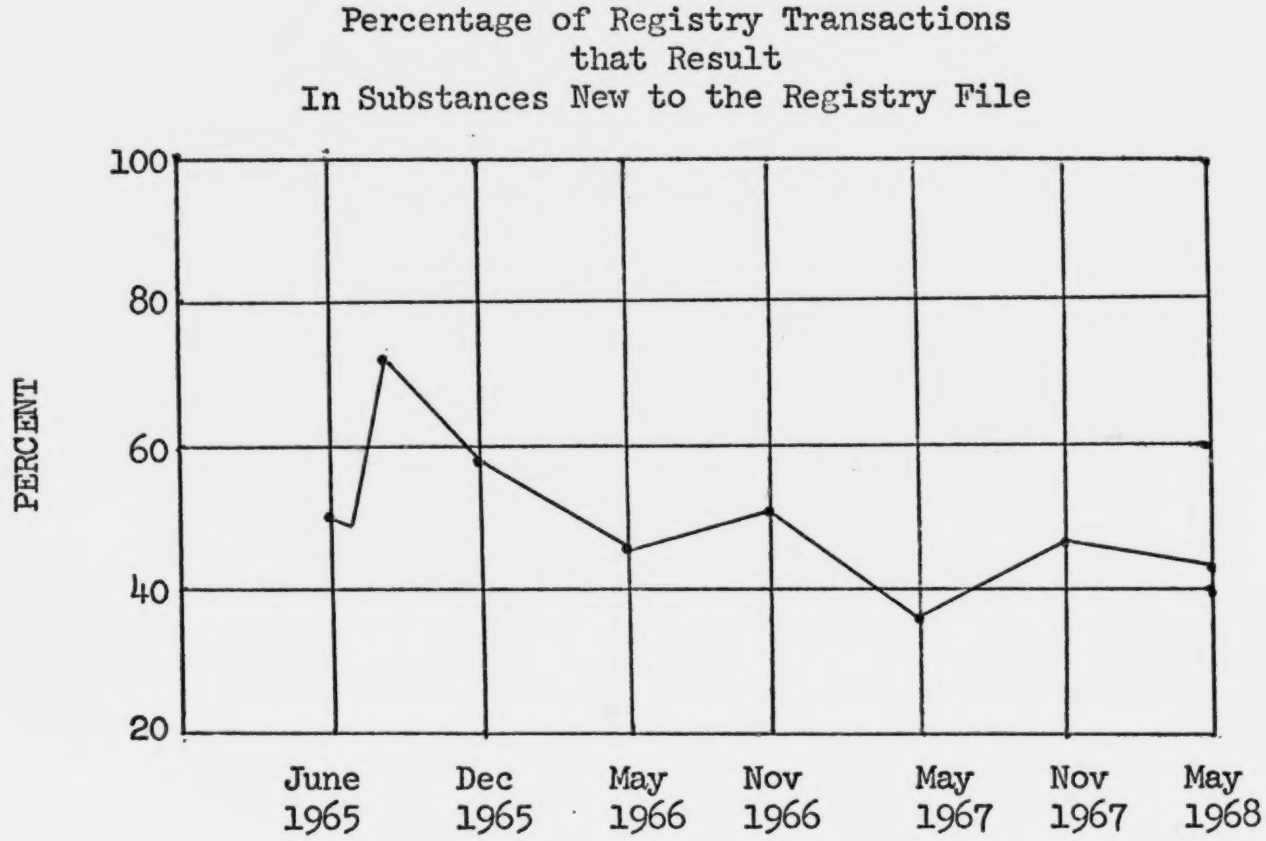
Percentage of Registry Transactions
that Result
In Substances New to the Registry File
PERCENT
100
80
60
40
20
June 1965
Dec 1965
May 1966
Nov 1966
May 1967
Nov 1967
May 1968

FIGURE I